



# High-Dimensional Range Profile Geometrical Visualization and Performance Estimation of Radar Target Classification via a Gaussian Mixture Model

Thomas Boulay, Ali Mohammad-Djafari, Nicolas Gac, Julien Lagoutte

## ► To cite this version:

Thomas Boulay, Ali Mohammad-Djafari, Nicolas Gac, Julien Lagoutte. High-Dimensional Range Profile Geometrical Visualization and Performance Estimation of Radar Target Classification via a Gaussian Mixture Model. Geometric Science of Information, Aug 2013, Paris, France. pp.829-836, 10.1007/978-3-642-40020-9\_93 . hal-00931023

**HAL Id: hal-00931023**

**<https://hal.science/hal-00931023>**

Submitted on 14 Jan 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# High-Dimensional Range profile geometrical visualization and performance estimation of radar target classification via a Gaussian mixture model

Thomas Boulay<sup>1</sup>, Ali Mohammad-Djafari<sup>2</sup>, Nicolas Gac<sup>2</sup>, and Julien Lagoutte<sup>1</sup>

<sup>1</sup> Surface Radar, Thales Air Systems, Limours, France,  
`thomas.boulay@thalesgroup.com`

<sup>2</sup> Laboratoire des Signaux et Systèmes,  
UMR 8506 CNRS-SUPELEC-UNIV PARIS SUD, Gif-sur-Yvette, France,  
`Ali.Mohammad-Djafari@lss.supelec.fr`

**Abstract.** In this paper, a method of data visualization and classification performance estimation applied to target classification is proposed. The objective of this paper is to propose a mathematical tool for data characterization. The principle is to use a non linear dimensionality reduction technique to describe our data in a low-dimensional space and to model embedding data by Gaussian mixture model (GMM) to estimate classification performance graphically and analytically.

## 1 Introduction and context

During major conflicts, cooperative classification techniques are not enough reliable enough and Non Cooperative Target Recognition (NCTR) [MO98] is increasingly seen as essential. High Range Resolution (HRR) [WE94] offers a rapid way to characterise a target through the use of radar range profile. A range profile is essentially a one-dimensional radar image of the target.

The HRR profile is a representation of temporal response of target to an high resolution radar impulse. Range resolution obtains with this kind of radar signature is less than a meter, which allows to take into account very small fluctuation of Radar Cross Section (RCS) along the target. As consideration, high resolution involves range profile with a large number of samples. Therefore, HRR profiles live in high-dimensional space. Most of classification techniques consist in comparing target signature under test with target signatures contained in a data set. In this case, classification problem is considered as a supervised classification problem [DU01]. Analysis of data set properties and visualization of classes repartition in this set can be interesting to estimate classification performance. To visualize high-dimensional data in a 2D or 3D space, a large number of dimensionality reduction techniques have been proposed for several years. Among these methods, t-Student Stochastic Neighbor Embedding [MA08] method allows to reduce the tendency to crowd points together in the center of a map

contrary to classical methods like Locally Linear Embedding [RO00], Isomap [TE00] or Sammon mapping [SA69].

This paper is organized as follows. In section 2, the database used for our study is described. The section 3 is a reminder on the dimensionality reduction methods and more particularly on the t-Distributed Stochastic Neighbor Embedding technique. In the section 4, a GMM algorithm for NCTR application is proposed. The section 5 shows the kind of data visualization obtained with our method. Probability map and decision matrix in low-dimensional space are computed and compared with decision matrix obtained in high-dimensional space with GMM algorithm. Our conclusion and suggestions for future work are presented in section 6.

## 2 Database

The data used in this article are synthetic data. The data set contains  $N_T$  range profiles in  $K = 3$  different classes ( $N_T^1 = N_T^2 = N_T^3 = 342$ ). They were generated from target modeling (CAD) on which were applied techniques to calculate RCS. For each configuration (frequency, elevation angle, azimuth), the RCS complex value is calculated. The table 1 summarizes notation used for the data set.

$M$	the number of range bins in our range profiles.
$N_T^k$	the number of range profiles belonging to the class $k$ .
$N_T = \sum_{k=1}^K N_T^k$	the total number of range profiles.
$\mathbf{w}_{T,i}^k$	the $i$ -th range profile belonging to the class $k$ , with $k = \{1, \dots, K\}$ et $i = [1, \dots, N_T^k]$ .
$\mathbf{x}_T^k$	the matrix of range profiles belonging to the class $k$ .
$\mathbf{w}_{T,i}$	the $i$ -th range profile with $i = [1, \dots, N_T]$ .
$\mathbf{x}_T$	the matrix of range profiles.
$\mathbf{y}_{T,i}^k$	the $i$ -th low-dimensional range profile belonging to the class $k$ .
$\mathbf{y}_T^k$	the matrix of low-dimensional range profiles belonging to the class $k$ .
$\mathbf{y}_{T,i}$	the $i$ -th low-dimensional range profile.
$\mathbf{y}_T$	the matrix of low-dimensional range profiles.

Table 1: Summary of notations used for the data set

## 3 Dimensionality reduction

A large number of nonlinear dimensionality reduction techniques that aims to preserve the local structure of data have been proposed for many years and many of which are reviewed by Lee and Verleysen [LE07]. Among these techniques, we can mention the most popular: Sammon Mapping [SA69], Stochastic Neighbor Embedding (SNE) [HI02], Isomap [TE00], Maximum Variance Unfolding

(MVU) [WE04], Locally Linear Embedding (LLE) [RO00] and Laplacian Eigenmaps [BE02]. On artificial data, these techniques obtain very good performance but they are not often very successful at visualizing real high-dimensional data. Indeed, most of these techniques are not capable of retaining both local and the global structure of the data in a single map.

In 2009, van der Maaten and Hinton [MA08] proposed a new method called “t-Distributed Stochastic Neighbor Embedding” or “t-SNE”, which is capable of capturing very well much of the local structure of the high-dimensional data, while also revealing global structure such as the presence of clusters at several scales. t-SNE method is directly inspired from SNE method.

### 3.1 Stochastic Neighbor Embedding (SNE)

Principle of SNE method is to convert high-dimensional Euclidian distances between datapoints into conditional probabilities that represent similarities. The similarity of range profile  $\mathbf{x}_{T,i}$  to range profile  $\mathbf{x}_{T,j}$  is the conditionnal probability,  $p_{j|i}$ , that  $\mathbf{x}_{T,i}$  would pick  $\mathbf{x}_{T,j}$  as its neighbor if neighbors were picked in proportion to their probability density under a Gaussian centered at  $\mathbf{x}_{T,i}$ . Mathematically, the conditional probability  $p_{j|i}$  is defined by

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_{T,i} - \mathbf{x}_{T,j}\|^2 / (2\sigma_i^2))}{\sum_{k \neq i} \exp(-\|\mathbf{x}_{T,i} - \mathbf{x}_{T,k}\|^2 / (2\sigma_i^2))} \quad (1)$$

where  $\sigma_i$  is the variance of the Gaussian that is centered on range profile  $\mathbf{x}_{T,i}$ .

For the low-dimensional range profiles  $\mathbf{y}_{T,i}$  and  $\mathbf{y}_{T,j}$  of the high-dimensional range profiles  $\mathbf{x}_{T,i}$  and  $\mathbf{x}_{T,j}$ , it is possible to compute a similar conditional probability, which is denote by  $q_{j|i}$

$$q_{j|i} = \frac{\exp(-\|\mathbf{y}_{T,i} - \mathbf{y}_{T,j}\|^2)}{\sum_{k \neq i} \exp(-\|\mathbf{y}_{T,i} - \mathbf{y}_{T,k}\|^2)} \quad (2)$$

If the map points  $\mathbf{y}_{T,i}$  and  $\mathbf{y}_{T,j}$  correctly model the similarity between the high-dimensional range profiles  $\mathbf{x}_{T,i}$  and  $\mathbf{x}_{T,j}$ , the conditional probabilities  $p_{j|i}$  and  $q_{j|i}$  will be equal. Therefore, SNE aims to find a low-dimensional data representation that minimizes the mismatch between  $p_{j|i}$  and  $q_{j|i}$ . The Kullback-Leiber divergence is a natural measure of the faithfulness with which  $q_{j|i}$  models  $p_{j|i}$ . SNE minimizes the sums of Kullback-Leiber divergences over all datapoints using a gradient descent method. The cost function  $C$  is given by

$$C = \sum_i KL(P_i || Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} \quad (3)$$

where  $P_i$  represents the conditional probability distribution over all other data points given datapoint  $\mathbf{x}_{T,i}$  and  $Q_i$  represents the conditional probability distribution over all other map points given map point  $\mathbf{y}_{T,i}$

### 3.2 t-Distributed Stochastic Neighbor Embedding (t-SNE)

t-SNE method provides two major innovations compared to classical SNE method. The cost function used by t-SNE differs from the one used by SNE in two ways. Firstly, it uses a symmetrized version of the SNE cost function with simpler gradients. Indeed, to minimizing the sum of the Kullback-Leiber divergences between the conditional probabilities  $p_{j|i}$  and  $q_{j|i}$ , it is also possible to minimize a single Kullback-Lieber divergence between a joint probability distribution,  $P$ , in the high-dimensional space and a joint probability distribution,  $Q$ , in the low-dimensional space:

$$C = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \text{ with } p_{ii} = p_{jj} = 0 \quad (4)$$

In symmetric SNE, the pairwise similarities in the low-dimensional map  $q_{ij}$  are given by

$$q_{ij} = \frac{\exp(-\|\mathbf{y}_{T,i} - \mathbf{y}_{T,j}\|^2)}{\sum_{k,l,k \neq l} \sum \exp(-\|\mathbf{y}_{T,k} - \mathbf{y}_{T,l}\|^2)} \quad (5)$$

and in the high-dimensionality space

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N_T} \quad (6)$$

This ensures that  $\sum_j p_{ij} > \frac{1}{2N_T}$  for all datapoints  $\mathbf{x}_{T,i}$ , as a result of which each datapoints  $\mathbf{x}_{T,i}$  makes a significant contribution to the cost function.

Secondly it uses the Student t-distribution rather than the Gaussian to compute the similarity between two points in the low-dimensional space. t-SNE employs a heavy-tailed distribution in the low-dimensional space to alleviate both the crowding problem and the optimization problems of SNE. Therefore, in t-SNE,  $q_{ij}$  becomes

$$q_{ij} = \frac{(1 + \|\mathbf{y}_{T,i} - \mathbf{y}_{T,j}\|^2)^{-1}}{\sum_{k,l,k \neq l} (1 + \|\mathbf{y}_{T,k} - \mathbf{y}_{T,l}\|^2)^{-1}} \quad (7)$$

## 4 Gaussian Mixture Models (GMM) for NCTR application

Let note by  $\mathbf{Z}_T$  the data set. Then for any of its element  $\mathbf{z}_{T,i}$  (in low-dimensional space  $\mathbf{z}_{T,i} = \mathbf{y}_{T,i} \in \mathbb{R}^2$  and in high-dimensional space  $\mathbf{z}_{T,i} = \mathbf{x}_{T,i} \in \mathbb{R}^M$ ), a mixture model can be defined as follows:

$$p(\mathbf{z}_{T,i}|\mathbf{Z}_T) = \sum_{k=1}^K \pi_T^k p(\mathbf{z}_{T,i}|\theta_T^k) \quad (8)$$

where  $\theta_T^k$  is the set of parameters of distribution  $k$  and  $\pi_T^k$  is the prior probability of the class  $k$ .

#### 4.1 Gaussian Mixture Models (GMM)

In the case of GMM,  $\theta_T^k = \{\mathbf{m}_T^k, \mathbf{C}_T^k\}$ , where  $\mathbf{m}_T^k$  ( $\mathbf{m}_T^k \in \mathbb{R}^2$  in the low-dimensional space and  $\mathbf{m}_T^k \in \mathbb{R}^M$  in the high-dimensional space) is the mean range profile of the class  $k$ :

$$\mathbf{m}_T^k = \frac{1}{N_T^k} \sum_{i=1}^{N_T^k} \mathbf{z}_{T,i}^k \quad (9)$$

and  $\mathbf{C}_T^k$  ( $\mathbf{C}_T^k \in \mathbb{R}^{2 \times 2}$  in the low-dimensional space and  $\mathbf{C}_T^k \in \mathbb{R}^{M \times M}$  in the high-dimensional space) is the covariance matrix of the class  $k$ . Elements of this matrix are defined by:

$$\mathbf{C}_T^k(p, q) = \frac{1}{N_T^k - 1} \sum_{i=1}^{N_T^k} (z_{T,i}^k(p) - m_T^k(p)) (z_{T,i}^k(q) - m_T^k(q)) \quad (10)$$

Finally, the Bayes rule is used to computed posterior probability of each class  $k$ , with  $p(\mathbf{z}_{T,i} | \theta_T^k) = \mathcal{N}(\mathbf{z}_{T,i} | \theta_T^k)$ .

$$p(\mathbf{z}_{T,i} \in k | \mathbf{z}_{T,i}) = \frac{\pi_T^k p(\mathbf{z}_{T,i} | \theta_T^k)}{\sum_i^K \pi_T^k p(\mathbf{z}_{T,i} | \theta_T^k)} \quad (11)$$

#### 4.2 Decision rule and error rate control

A major constraint in NCTR application is to control error rate to avoid fratricide shoots. In this context, the classical decision rule of maximum a posteriori (MAP) cannot be used. We defined a new decision rule adapted to NCTR context. Let  $\mathbf{p} = [p_1, \dots, p_K]$  be the vector of posterior probability computed for the  $K$  classes.

- if  $p_i \geq TG$ , the class  $i$  is granted
- if  $TD \leq p_i < TG$ , the class  $i$  is dubious
- if  $p_i < TD$ , le class  $i$  is denied

We construct three decision matrix: (i) a Granted matrix (**Grm**), (ii) a Dubious matrix (**Dum**) and (iii) a Denied matrix (**Dem**). Each of them is a  $K \times K$  matrix. The decision thresholds  $TG$  and  $TD$  are ajusted empirically to maintain error rate <sup>3</sup> less than a fixed value (for example 5%).

<sup>3</sup> (i) We have an error when the class  $i$  is declared denied for a range profile of the class  $i$ . Diagonal of **Dem** gives the error rate. (ii) We have a success when the class  $i$  is declared granted and the other classes dubious or denied for a range profile of the class  $i$ . Success rate ( $SR$ ) is deduced from  $Grm$ . ( $SR(i) = Grm(i) \times \prod_{j \neq i} (1 - Grm(j))$ ). (iii) We have a good identification when the class  $i$  is declared granted and the others classes denied for a range profile of the class  $i$ . Good identification rate ( $GIR$ ) is deduced from  $Grm$  and  $Dem$ . ( $GIR(i) = Grm(i) \times \prod_{j \neq i} Dem(j)$ ).

## 5 Data visualization and classification performance estimation

### 5.1 Probability map in low-dimensional space

After computing  $\mathbf{Y}_T$  from t-SNE algorithm and from  $\mathbf{X}_T$ , we want to visualize how data are placed over the decision rule. First from  $\mathbf{Y}_T$ , we estimate parameters  $\theta_T^k$  of the  $K$  Gaussian distribution of GMM. Then, we can sample the 2D space and calculate for each sample of the grid the posterior probability for each Gaussian distribution of GMM. Finally, we can plot probability maps where granted, dubious and denied zones for each class are represented. The low-dimensional range profile can be superposed on this map and we can estimate graphically classification performance with our data.

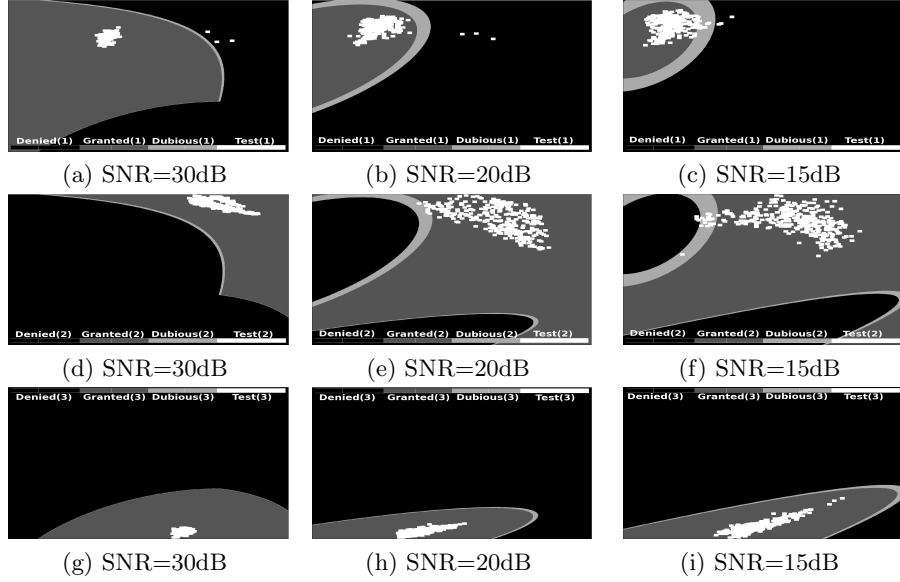


Fig. 1: t-SNE 2D visualization of data for class 1, 2 and 3 (row) and for three SNR values (column).

From figure 1a, 1d and 1g, we can see immediately that error rate for class 2 and 3 will be equal to 0% and good identification rate equal to 100%, because each sample of class 2 and 3 is entirely in granted area. For the class 1, we can see that some samples are in class 1 denied area and in class 2 granted area. With  $\text{SNR} = 20\text{dB}$  (cf. figure 1b, 1e and 1h), some samples of class 1 are yet placed in class 2 granted area and some samples of class 2 are in class 1 dubious area. Samples of class 3 remain in class 3 granted area. When  $\text{SNR} = 15\text{dB}$  (cf. figure 1c, 1f and 1i), increasingly samples of class 1 and 2 are placed in class 1 dubious area.

## 5.2 Decision matrix in low-dimensional space

We can retrieve graphically observations by computing **Grm**, **Dum** and **Dem** matrix defined in section 4.2.

	Test \ Class	Grm			Dum			Dem		
		1	2	3	1	2	3	1	2	3
SNR=30dB	Class 1	99.1	0.9	0	0	0	0	0.9	99.1	100
	Class 2	0	100	0	0	0	0	100	0	100
	Class 3	0	0	100	0	0	0	100	100	0
SNR=20dB	Class 1	99.1	0.9	0	0	0	0	0.9	99.1	100
	Class 2	0	98.5	0	0.9	1.5	0	98.5	0	100
	Class 3	0	0	100	0	0	0	100	100	0
SNR=15dB	Class 1	95.6	0.9	0	3.5	1.1	0	0.9	98	100
	Class 2	0.6	94.4	0	5	4.7	0	94.4	0.9	100
	Class 3	0	0	100	0	0	0	100	100	0

Fig. 2: **Grm**, **Dum** and **Dem** matrix for three different SNR in low-dimensional space

Indeed, as seen on the probability maps, error rate and good identification rate (and success rate) are respectively equal to 0% and 100% for class 2 and 3 with an SNR = 30dB. With SNR = 30dB, SNR=20dB or SNR=15dB, we retrieve that some samples (0.9%) of class 1 are placed in class 2 granted area. When SNR = 15dB, some samples (3.5%) of class 1 are placed in class 1 dubious area and in class 2 dubious area (1.1%). Generally, this table allows us to get an analytical measurement of the observations made from probability maps.

## 5.3 Decision matrix in high-dimensional space

	Test \ Class	Grm			Dum			Dem		
		1	2	3	1	2	3	1	2	3
SNR=30dB	Class 1	100	0	0	0	0	0	0	100	100
	Class 2	0	100	0	0	0	0	100	0	100
	Class 3	0	0	100	0	0	0	100	100	0
SNR=20dB	Class 1	99.4	0.6	0	0	0.3	0	0.6	99.1	100
	Class 2	0.3	99.7	0	0	0.3	0	99.7	0	100
	Class 3	0	0	100	0	0	0	100	100	0
SNR=15dB	Class 1	97.7	2.3	0	0.3	0	0	2	97.7	100
	Class 2	0.3	99.7	0	0	0	0	99.7	0.3	100
	Class 3	0	0	100	0	0	0	100	100	0

Fig. 3: **Grm**, **Dum** and **Dem** matrix for three different SNR in high-dimensional space



Finally, we can compute **Grm**, **Dum** and **Dem** matrix obtained applying GMM algorithm in the high-dimensional space (cf. Fig.3). Overall, classification performances are better in high-dimensional space than in low-dimensional space. Even if rates are higher in high-dimensional space, we retrieve the same tendency than we can observe in low-dimensional space. Our method provide a good tool to estimate analytically and graphically classification performance we can get from a data set and its associated probability distribution model.

## 6 Conclusion and perspectives

In this paper, a new method of classification performance evaluation is proposed. This method mixes dimensionality reduction and classical Gaussian mixture modelization to propose a mathematical tool for classification performance evaluation. An additional work will be to generalize these methods for any kind of algorithm and to evaluate for each kind of algorithm the critical dimension from which the performances begin to degrade. This will allow to determine the best appropriate algorithm from data and to reduce the computational cost of this algorithm.

## References

- [MA08] van der Maaten, L., Hinton, G.: Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9** (2008) 2579–2605
- [WE94] Wehner, R., Barnes, B.: *High Resolution Radar*. Artech House Publishers (1994)
- [DU01] Duda, Richard O., Hart, Peter E., Stork, David G.: *Pattern Classification*. Wiley Interscience (2001)
- [LE07] Lee, J. A., Verleysen, M.: *Nonlinear dimensionality reduction*. Springer, New York, NY, USA (2007)
- [RO00] Roweis, S. T., Saul, L. K.: Nonlinear dimensionality reduction by Locally Linear Embedding. *Science*, 290(5500) (2000) 2323–2326
- [HI02] Hinton, G. E., Roweis, S. T.: Stochastic Neighbor Embedding. In *Advances in Neural Information Processing Systems*, Cambridge (USA), Volume 15, (2002) 833–840
- [BE02] Belkin, M., Nigoyi, P.: Laplacian Eignemaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems*, Cambridge (USA), Volume 14, (2002) 585–591
- [WE04] Weinberger, K. Q., Sha, F., Saul, L. K.: Learning a kernel matrix for non linear dimensionality reduction. In *Proceedings of the International Conference on Machine Learning*, (2004)
- [SA69] Sammon, J. W.: A nonlinear mapping for data structure analysis. *IEEE Transactions on Computer*, 18(5) (1969) 401–409
- [TE00] Tenenbaum, J. B., de Silva, V., Langford, J. C.: A global geometric framework for non linear dimensionality reduction. *Science*, 290(5500) (2000) 2319–2323
- [MO98] Moruzzis, M., Colin, N.: Radar target recognition by Fuzzy Logic. *IEEE Aerospace and Electronic Systems Magazine*, Volume 13 (1998) 13–20